Feature-based Time Series Forecasting

Thiyanga Talagala,
Rob J Hyndman, George Athanasopoulos,
Feng Li, Yanfei Kang

11 July 2019
What algorithm is likely to perform best?
What algorithm is likely to perform best?
Algorithm selection problem, John Rice (1976)
Time series features

- Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
Time series features

- Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'$.
More features

- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- spectral entropy
- Hurst exponent
- nonlinearity
- unit root test statistics
- parameter estimates of Holt’s linear trend method
- parameter estimates of Holt-Winters’ additive method
- ACF and PACF based features - calculated on raw, differenced, seasonally-differenced series and remainder series.
Algorithm selection framework
Algorithm selection framework

population \rightarrow sample
Algorithm selection framework

- population
- sample
- training set
- test set
Algorithm selection framework

- Population
- Sample
- Training set
- Input feature
- Test set
Algorithm selection framework
Algorithm selection framework

population → sample

training set → input - feature → fit models

test set

<table>
<thead>
<tr>
<th>id</th>
<th>arima</th>
<th>ets</th>
<th>nn</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.23</td>
<td>1.01</td>
<td>2.51</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>3.51</td>
<td>4.51</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>0.14</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Algorithm performance space
Algorithm selection framework
Algorithm selection framework

- **population**
- **sample**
- **training set**
  - **input - feature**
  - **fit models**
    - **test set**
    - **algorithm performance space**
    - **train a meta-learner**
Algorithm selection framework

- **Population** → **Sample**
- **Training set** → **Input feature** → **Fit models**
- **Test set**
- **Algorithm performance space**
- **Train a meta-learner**

<table>
<thead>
<tr>
<th>id</th>
<th>arima</th>
<th>ets</th>
<th>nn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.23</td>
<td>1.01</td>
<td>2.51</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>3.51</td>
<td>4.51</td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>0.14</td>
<td>0.50</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Algorithm selection framework

**Population** → **Sample** → **Training Set** → **Input Feature** → **Fit Models** → **Test Set** → **Algorithm Performance Space** → **Train a Meta-Learner** → **Meta-Learner** → **New Time Series**
Algorithm selection framework

Population -> Sample -> Training set -> Input - Feature -> Fit Models -> Test set

New time series -> Feature calculation

Train a meta-learner

<table>
<thead>
<tr>
<th>id</th>
<th>arima</th>
<th>ets</th>
<th>nn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.23</td>
<td>1.01</td>
<td>2.51</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>3.51</td>
<td>4.51</td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>0.14</td>
<td>0.50</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Algorithm performance space
Algorithm selection framework

population -> sample -> training set -> input - feature -> fit models -> test set

new time series -> feature calculation

train a meta-learner

Algorithm performance space

Table:

<table>
<thead>
<tr>
<th>id</th>
<th>arima</th>
<th>ets</th>
<th>nn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.23</td>
<td>1.01</td>
<td>2.51</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>3.51</td>
<td>4.51</td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>0.14</td>
<td>0.50</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Algorithm selection framework

new time series → feature calculation → meta-learner

population → sample → training set → input - feature → fit models → test set

Algorithm performance space

id  arima  ets  nn  ...
1   1.23   1.01  2.51
2   0.06   3.51  4.51
3   0.80   0.14  0.50
...

train a meta-learner

forecasting method
Algorithm selection framework

1. **population** → **sample**
2. **training set** → **input - feature** → **fit models**
3. **test set**
4. **new time series** → **feature calculation**
5. **Algorithm performance space**
6. **train a meta-learner**
7. **meta-learner** → **forecasting method**
FFORMS: Feature-based FORecast Model Selection

- two algorithms: FFORMS, FFORMPP
FFORMPP: Feature-based FORecast Model Performance Prediction

- two algorithms: FFORMS, FFORMPP
seer R package

Installation

```r
devtools::install_github("thiyangt/seer")
library(seer)
```
seer R package

Installation

devtools::install_github("thiyangt/seer")
library(seer)

Example dataset

observed time series - M1 yearly series (181)

library(Mcomp)
yearlyM1 <- subset(M1, "yearly")
seer R package

Installation

devtools::install_github("thiyangt/seer")
library(seer)

Example dataset

observed time series - M1 yearly series (181)

library(Mcomp)
yearlyM1 <- subset(M1, "yearly")
Input: features

cal_features(yearlym1[1:2], database="M1", h=6, highfreq=FALSE)

# A tibble: 2 x 25
  entropy lumpiness stability hurst trend spikiness linearity curvature
             <dbl>       <dbl>      <dbl> <dbl> <dbl>  <dbl>     <dbl>     <dbl>
1       0.683   0.0400     0.977 0.985 0.985 1.32e-6  4.46     0.705
2       0.711   0.0790     0.894 0.988 0.989 1.54e-6  4.47     0.613
# ... with 17 more variables: e_acf1 <dbl>, y_acf1 <dbl>,
# diff1y_acf1 <dbl>, diff2y_acf1 <dbl>, y_pacf5 <dbl>,
# diff1y_pacf5 <dbl>, diff2y_pacf5 <dbl>, nonlinearity <dbl>,
# lmres_acf1 <dbl>, ur_pp <dbl>, ur_kpss <dbl>, N <int>, y_acf5 <dbl>,
# diff1y_acf5 <dbl>, diff2y_acf5 <dbl>, alpha <dbl>, beta <dbl>
seer::fcast_accuracy(tsls=yearlym1[1:2],
models=c("arima","ets","rw","theta","nn"),
database="M1", cal_MASE, h=6,
length_out=1,
fcast_save=TRUE)

$accuracy
              arima   ets    rw    theta     nn
YAF2  10.527612 10.319029 13.52428 12.088375 11.78891
YAF3    5.713867  7.704409  7.78949  6.225463  6.70074

$ARIMA
YAF2          YAF3
"ARIMA(0,1,0) with drift"  "ARIMA(0,1,1) with drift"

$ETS
YAF2          YAF3
"ETS(A,A,N)"  "ETS(M,A,N)"

$forecasts
$forecasts$arima
YAF2          YAF3
 [1,]  579581.0  390955.9
 [2,]  605761.9  407325.1
 [3,]  631942.9  423694.4
 [4,]  658123.8  440063.6
 [5,]  684304.8  456432.8
 [6,]  710485.7  472802.0

$forecasts$ets
YAF2          YAF3
 [1,]  556280.7  384603.9
 [2,]  594333.0  385162.7
 [3,]  632385.3  385721.5
 [4,]  670437.6  386280.4
 [5,]  708489.9  386839.2
 [6,]  746542.3  387398.0

$forecasts$rw
YAF2          YAF3
 [1,]  553400.0  384040.0
 [2,]  553400.0  384040.0
 [3,]  553400.0  384040.0
 [4,]  553400.0  384040.0
 [5,]  553400.0  384040.0
 [6,]  553400.0  384040.0

$forecasts$theta
YAF2          YAF3
 [1,]  565938.8  394342.6
 [2,]  578486.4  404640.6
 [3,]  591034.0  414938.7
 [4,]  603581.5  425236.7
 [5,]  616129.1  435534.8
 [6,]  628676.6  445832.8

$forecasts$nn
YAF2          YAF3
 [1,]  575702.1  399629.8
 [2,]  592979.1  407195.2
 [3,]  605912.6  410557.7
 [4,]  615336.6  411990.4
 [5,]  622064.3  412589.6
 [6,]  626795.5  412838.3
```
seer::fcast_accuracy(tslist=yearlyM1[1:2],
  models= c("arima","ets","rw", "theta", "nn"),
  database ="M1", cal_MASE, h=6,
  length_out = 1,
  fcast_save = TRUE)
```

$accuracy

<table>
<thead>
<tr>
<th></th>
<th>arima</th>
<th>ets</th>
<th>rw</th>
<th>theta</th>
<th>nn</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAF2</td>
<td>10.527612</td>
<td>10.319029</td>
<td>13.52428</td>
<td>12.088375</td>
<td>11.78891</td>
</tr>
<tr>
<td>YAF3</td>
<td>5.713867</td>
<td>7.704409</td>
<td>7.78949</td>
<td>6.225463</td>
<td>6.70074</td>
</tr>
</tbody>
</table>

$ARIMA

YAF2       YAF3
"ARIMA(0,1,0) with drift" "ARIMA(0,1,1) with drift"

$ETS

YAF2       YAF3
"ETS(A,A,N)" "ETS(M,A,N)"

$forecasts

$forecasts$arima

YAF2       YAF3
[1,] 579581.0 390955.9
[2,] 605761.9 407325.1
[3,] 631942.9 423694.4
[4,] 658123.8 440063.6
[5,] 684304.8 456432.8
[6,] 710485.7 472802.0

$forecasts$ets

YAF2       YAF3
[1,] 556280.7 384603.9
[2,] 594333.0 385162.7
[3,] 632385.3 385721.5
[4,] 670437.6 386280.4
[5,] 708489.9 386839.2
[6,] 746542.3 387398.0

$forecasts$rw

YAF2       YAF3
[1,] 553400   384040
[2,] 553400   384040
[3,] 553400   384040
[4,] 553400   384040
[5,] 553400   384040
[6,] 553400   384040

$forecasts$theta

YAF2       YAF3
[1,] 565938.8 394342.6
[2,] 578486.4 404640.6
[3,] 591034.0 414938.7
[4,] 603581.5 425236.7
[5,] 616129.1 435534.8
[6,] 628676.6 445832.8

$forecasts$nn

YAF2       YAF3
[1,] 575702.1 399629.8
[2,] 592979.1 407195.2
[3,] 605912.6 410557.7
[4,] 615336.6 411990.4
[5,] 622064.3 412589.6
[6,] 626795.5 412838.3

MASE

\[
q_t = \frac{1}{n-1} \sum_{i=2}^{n} |Y_i - Y_{i-1}|
\]

\[
MASE = \text{mean}(|q_t|)
\]
prepare_trainingset(accuracy_set = accuracy_m1, feature_set = features_m1)$trainingset

# A tibble: 2 x 26
                 entropy    lumpiness   stability   hurst   trend  spikiness linearity curvature
          <dbl>      <dbl>       <dbl>     <dbl>   <dbl>    <dbl>     <dbl>      <dbl>
1   0.6832709 0.039980970 0.97700493 0.9850887 1.32e-06 0.00446574 0.7051875
2   0.7110281 0.079043840 0.89417894 0.9880270 1.54e-06 0.00446947 0.6132081

# ... with 18 more variables: e_acf1 <dbl>, y_acf1 <dbl>,
#   diff1y_acf1 <dbl>, diff2y_acf1 <dbl>, y_pacf5 <dbl>,
#   diff1y_pacf5 <dbl>, diff2y_pacf5 <dbl>, nonlinearity <dbl>,
#   lmres_acf1 <dbl>, ur_pp <dbl>, ur_kpss <dbl>, N <int>, y_acf5 <dbl>,
#   diff1y_acf5 <dbl>, diff2y_acf5 <dbl>, alpha <dbl>, beta <dbl>,
#   classlabels <chr>
FFORMS classifier

```r
rf <- build_rf(training_set = training_set, 
                testset= M3yearly_features, 
                rf_type="ru", ntree=100, seed=1, 
                import=FALSE, mtry = 8)

Predictions

head(rf$predictions)
```

```
## 1 2 3 4 5 6
## ETS-trend rwd rwd rwd rwd rwd
## 10 Levels: ARIMA ARMA/AR/MA ETS-dampedtrend ... wn
```

FFORMS classifier

```r
rf$randomforest
```

```
## randomForest(formula = classlabels ~ ., data = training_set, 
## importance = import, ntree = ntree, mtry = mtry)
```
Pre-trained classifiers

Load FFORMS classifier for hourly series

data("hourly_fforms")
Pre-trained classifiers

Load FFORMS classifier for hourly series

```r
data("hourly_fforms")
```

Forecast hourly time series in the M4-competition

```r
fcast.models <- predict(hourly_fforms, features_M4H)
head(fcast.models)
```

```
## 1 2 3 4 5 6
## Levels: mstlarima mstlets nn rw rwd snaive stlar tbats theta wn
```
Yearly: Correlation between MASE values across different forecast-models
**FFORMPP: Feature-based FORecast Model Performance Prediction**

- Efficient Bayesian Multivariate Surface Regression (Feng Li & Mattias Villani, 2013)
  - handles interactions and nonlinear relationships
  - allows the knot locations to move freely in the feature space

![Table](image)

### Table

<table>
<thead>
<tr>
<th>id</th>
<th>seasonality</th>
<th>...</th>
<th>entropy</th>
<th>trend</th>
<th>rw</th>
<th>rwd</th>
<th>....</th>
<th>arima</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.89</td>
<td></td>
<td>0.21</td>
<td>0.82</td>
<td>1.02</td>
<td>0.89</td>
<td></td>
<td>0.78</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td></td>
<td>0.82</td>
<td>0.10</td>
<td>1.10</td>
<td>2.81</td>
<td></td>
<td>2.87</td>
</tr>
<tr>
<td>N</td>
<td>0.50</td>
<td></td>
<td>0.40</td>
<td>0.30</td>
<td>0.87</td>
<td>0.89</td>
<td></td>
<td>0.99</td>
</tr>
</tbody>
</table>

X-features  Y-MASE
fformpp R package

Installation

```r
devtools::install_github("thiyangt/fformpp")
library(fformpp)
```

Train a model

```r
fit_fformpp(feamat=features_mat, accmat=forecast.error,
sknots=2, aknots=2,
fix.s=0, fix.a=0, fix.shrinkage=1:5,
fix.covariance=0,
fix.coefficients=0, n.iter=100,
knot.moving.algorithm="Random-Walk",
ptype=c("identity", "identity", "identity"),
prior.knots=100)
```
predict.m1 <- predict(fformpp.model, features.m1.df, 
  c("ets", "arima", "rw", "rwd", "wn", "theta", "nn"), 
  log=FALSE, final.estimate=median)

head(predict.m1)

##    ets   arima    rw   rwd    wn   theta   nn
## [1,] 5.015336 5.065616 5.149868 4.293450 16.681046 4.316341 4.554838
## [2,] 1.990880 1.831033 1.830689 2.010443 7.845106 1.434183 2.864783
## [4,] 2.169089 3.162256 2.178721 2.481028 3.126736 2.216428 1.832553
## [5,] 5.199962 3.970234 4.630903 4.174412 15.631346 4.101041 5.765485
## Results: M4 Competition data

<table>
<thead>
<tr>
<th>Method</th>
<th>Yearly</th>
<th>Quarterly</th>
<th>Monthly</th>
<th>Weekly</th>
<th>Daily</th>
<th>Hourly</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFORMS_individual</td>
<td>3.17</td>
<td>1.20</td>
<td>0.98</td>
<td>2.31</td>
<td>3.57</td>
<td>0.84</td>
</tr>
<tr>
<td>FFORMPP_combination</td>
<td>3.07</td>
<td>1.13</td>
<td>0.89</td>
<td>2.46</td>
<td>3.62</td>
<td>0.96</td>
</tr>
<tr>
<td>auto.arima</td>
<td>3.40</td>
<td>1.17</td>
<td>0.93</td>
<td>2.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ets</td>
<td>3.44</td>
<td>1.16</td>
<td>0.95</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>theta</td>
<td>3.37</td>
<td>1.24</td>
<td>0.97</td>
<td>2.64</td>
<td>3.33</td>
<td>1.59</td>
</tr>
<tr>
<td>rwd</td>
<td>3.07</td>
<td>1.33</td>
<td>1.18</td>
<td>2.68</td>
<td>3.25</td>
<td>11.45</td>
</tr>
<tr>
<td>rw</td>
<td>3.97</td>
<td>1.48</td>
<td>1.21</td>
<td>2.78</td>
<td>3.27</td>
<td>11.60</td>
</tr>
<tr>
<td>nn</td>
<td>4.06</td>
<td>1.55</td>
<td>1.14</td>
<td>4.04</td>
<td>3.90</td>
<td>1.09</td>
</tr>
<tr>
<td>stlar</td>
<td>-</td>
<td>2.02</td>
<td>1.33</td>
<td>3.15</td>
<td>4.49</td>
<td>1.49</td>
</tr>
<tr>
<td>snaive</td>
<td>-</td>
<td>1.66</td>
<td>1.26</td>
<td>2.78</td>
<td>24.46</td>
<td>2.86</td>
</tr>
<tr>
<td>tbats</td>
<td>-</td>
<td>1.19</td>
<td>1.05</td>
<td>2.49</td>
<td>3.27</td>
<td>1.30</td>
</tr>
<tr>
<td>wn</td>
<td>13.42</td>
<td>6.50</td>
<td>4.11</td>
<td>49.91</td>
<td>38.07</td>
<td>11.68</td>
</tr>
<tr>
<td>mstlarima</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.84</td>
<td>1.12</td>
</tr>
<tr>
<td>mstlets</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.73</td>
<td>1.23</td>
</tr>
<tr>
<td>combination (mean)</td>
<td>4.09</td>
<td>1.58</td>
<td>1.16</td>
<td>6.96</td>
<td>7.94</td>
<td>3.93</td>
</tr>
<tr>
<td>M4-1st</td>
<td>2.98</td>
<td>1.12</td>
<td>0.88</td>
<td>2.36</td>
<td>3.45</td>
<td>0.89</td>
</tr>
<tr>
<td>M4-2nd</td>
<td>3.06</td>
<td>1.11</td>
<td>0.89</td>
<td>2.11</td>
<td>3.34</td>
<td>0.81</td>
</tr>
<tr>
<td>M4-3rd</td>
<td>3.13</td>
<td>1.23</td>
<td>0.95</td>
<td>2.16</td>
<td>2.64</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Thank you

R packages and papers

R packages
- seer: FFORMS
  github.com/thiyangt/seer
- fformpp: FFORMPP
  github.com/thiyangt/fformpp

Papers and Slides
thiyanga.netlify.com/talk/user19-talk/

email: thiyanga.talagala@monash.edu